Billboard100

Jason Antal

2024-08-18

Part A: Make a table of the top 10 most popular songs since 1958, as measured by the total number of weeks that a song spent on the Billboard Top 100 chart since 1958, up through the middle of 2021.Your table should have 10 rows and 3 columns: performer, song, and count, where count represents the number of weeks that song appeared in the Billboard Top 100. Make sure the entries are sorted in descending order of the count variable, so that the more popular songs appear at the top of the table. Give your table a short caption describing what is shown in the table.

```
top_10_songs <- billboard %>%
  group_by(performer, song) %>%
  summarise(count = n(), .groups = "drop") %>%
  arrange(desc(count)) %>%
  head(10)

print("The top 10 Most Popular Songs on the Billboard Hot 100 (Jan 1958- May 2021), ranked by total num
```

```
## [1] "The top 10 Most Popular Songs on the Billboard Hot 100 (Jan 1958- May 2021), ranked by total num
```

```
print(top_10_songs)
```

```
## # A tibble: 10 x 3
##    performer                                song                          count
##    <chr>                                    <chr>                         <int>
##  1 Imagine Dragons                          Radioactive                      87
##  2 AWOLNATION                               Sail                             79
##  3 Jason Mraz                               I'm Yours                        76
##  4 The Weeknd                               Blinding Lights                  76
##  5 LeAnn Rimes                              How Do I Live                    69
##  6 LMFAO Featuring Lauren Bennett & GoonRock Party Rock Anthem              68
##  7 OneRepublic                              Counting Stars                   68
##  8 Adele                                    Rolling In The Deep              65
##  9 Jewel                                    Foolish Games/You Were Meant~    65
## 10 Carrie Underwood                         Before He Cheats                 64
```

Part B: Is the "musical diversity" of the Billboard Top 100 changing over time? Let's find out. We'll measure the musical diversity of given year as the number of unique songs that appeared in the Billboard Top 100 that year. Make a line graph that plots this measure of musical diversity over the years. The x axis should show the year, while the y axis should show the number of unique songs appearing at any position on the Billboard Top 100 chart in any week that year. For this part, please filter the data set so that it excludes the years 1958 and 2021, since we do not have complete data on either of those years. Give the figure an informative caption in which you explain what is shown in the figure and comment on any interesting trends you see.

There are number of ways to accomplish the data wrangling here. For example, you could use two distinct sets of data-wrangling steps. The first set of steps would get you a table that counts the number of times that a given song appears on the Top 100 in a given year. The second set of steps operate on the result of the first set of steps; it would count the number of unique songs that appeared on the Top 100 in each year, irrespective of how many times it had appeared.
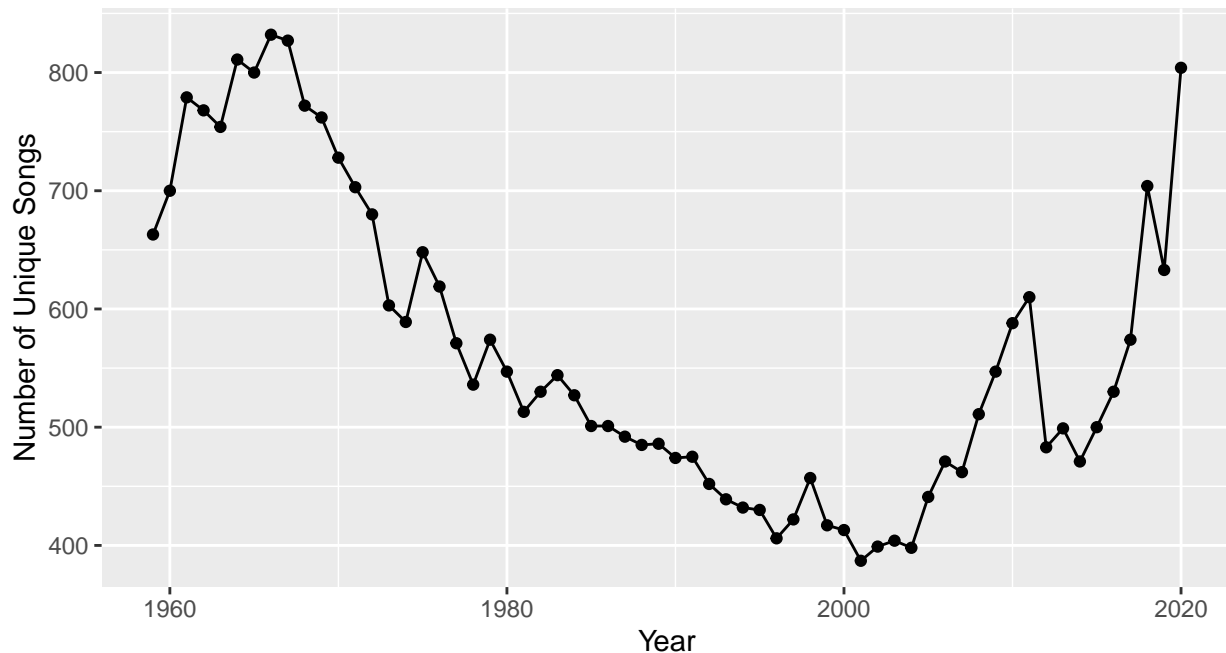
```r
#How many unique songs appeared in the Billboard 100 every year from '59-'20?

musical_diversity <- billboard %>%
  filter(year > 1958 & year < 2021) %>%
  distinct(year, song, performer) %>%
  group_by(year) %>%
  summarise(unique_songs = n(), .groups = "drop")

# Create the line graph
ggplot(musical_diversity, aes(x = year, y = unique_songs)) +
  geom_line() +
  geom_point() +
  labs(
    title = "Musical Diversity of Billboard Top 100 (1959-2020)",
    x = "Year",
    y = "Number of Unique Songs",
    caption = "A graph showing the number of unique songs from 1959-2020. At the beginning of this chart
    diversity in the billboard 100 was at an all-time high in the 1960's, with over 800 unique songs fro
    diversity then fell every year afterwards almost without exception, reaching its lowest point in 200
    unique songs. We then experienced a sharp rise in unique songs in the mid-late 2000's, a sharp fall
    another dramatic rise from 2015-2020 that ended with 2020 breaking the 800-unique-songs barrier aga
  )
```

Musical Diversity of Billboard Top 100 (1959–2020)

graph showing the number of unique songs from 1959–2020. At the beginning of this chart, we can see that musical rsity in the billboard 100 was at an all–time high in the 1960's, with over 800 unique songs from 1964–1967. Musical diversity then fell every year afterwards almost without exception, reaching its lowest point in 2001 at less than 400 unique songs. We then experienced a sharp rise in unique songs in the mid–late 2000's, a sharp fall in 2012, and another dramatic rise from 2015–2020 that ended with 2020 breaking the 800–unique–songs barrier again.

Part C: Let's define a "ten-week hit" as a single song that appeared on the Billboard Top 100 for at least ten weeks. There are 19 artists in U.S. musical history since 1958 who have had at least 30 songs that were "ten-week hits." Make a bar plot for these 19 artists, showing how many ten-week hits each one had in their musical career. Give the plot an informative caption in which you explain what is shown.

```r
ten_week_hits <- billboard %>%
  group_by(performer, song) %>%
  summarise(weeks_on_chart = n(), .groups = "drop") %>%
  filter(weeks_on_chart >= 10) %>%
  group_by(performer) %>%
  summarise(ten_week_hits = n()) %>%
  filter(ten_week_hits >= 30)

ggplot(ten_week_hits, aes(x = reorder(performer, ten_week_hits), y = ten_week_hits)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  coord_flip() +  # Flip coordinates for horizontal bars
  labs(
    title = "Artists with 30+ Ten-Week Hits on Billboard Top 100 (1958-2021)",
    x = "Artist",
    y = "Number of Ten-Week Hits",
    caption = 'placeholder')
```

Artists with 30+ Ten−Week Hits on Billboard Top 100 (1958−2(

placeholder